

Artificial intelligence models for predicting the performance of biological wastewater treatment plant in the removal of Kjeldahl Nitrogen from wastewater

D. S. Manu¹  · Arun Kumar Thalla¹

Received: 9 February 2016 / Accepted: 3 January 2017 / Published online: 13 January 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract The current work demonstrates the support vector machine (SVM) and adaptive neuro-fuzzy inference system (ANFIS) modeling to assess the removal efficiency of Kjeldahl Nitrogen of a full-scale aerobic biological wastewater treatment plant. The influent variables such as pH, chemical oxygen demand, total solids (TS), free ammonia, ammonia nitrogen and Kjeldahl Nitrogen are used as input variables during modeling. Model development focused on postulating an adaptive, functional, real-time and alternative approach for modeling the removal efficiency of Kjeldahl Nitrogen. The input variables used for modeling were daily time series data recorded at wastewater treatment plant (WWTP) located in Mangalore during the period June 2014–September 2014. The performance of ANFIS model developed using Gbell and trapezoidal membership functions (MFs) and SVM are assessed using different statistical indices like root mean square error, correlation coefficients (CC) and Nash Sutcliffe error (NSE). The errors related to the prediction of effluent Kjeldahl Nitrogen concentration by the SVM modeling appeared to be reasonable when compared to that of ANFIS models with Gbell and trapezoidal MF. From the performance evaluation of the developed SVM model, it is observed that the approach is capable to define the inter-relationship between various wastewater quality variables and thus SVM can be potentially applied for evaluating the efficiency of aerobic biological processes in WWTP.

Keywords ANFIS · SVM · Statistical indices · Total Kjeldahl Nitrogen · Wastewater treatment plants · Membership function

Introduction

Improper maintenance of WWTP can trigger serious ecological and public health problems and also it may be a reason for various water borne diseases affecting human health and aquatic life. Nitrogen and phosphorous are the key nutrients supporting the growth of algae and organic matter which instigate eutrophication in water bodies. Various control actions have to be implemented for efficient monitoring of process performance during the operation of wastewater treatment plant (WWTP) (Boelee et al. 2011). Models are necessary for the reason that, the effects of tuning the operating variables can be studied more transiently on a computer than by doing experiments. Hence, many alternative schemes and operational strategies can be evaluated without the need for physical trials of each scenario (Thalla et al. 2010; Pai et al. 2011). By simulating the performance assessment models using suitable influential variables, one can rapidly respond to any changes in the processes and devise operational strategies to shift the plant to new operating conditions which improves its stability, the quality of the effluent and at the same time achieve reduction in the running costs (Miller et al. 1997; Nair et al. 2016; Kumar and Saravanan 2009). Several deterministic, stochastic and time series-based models have been developed for predicting the performance of WWTPs, (Guo et al. 2014; Ráduly et al. 2007; Denai et al. 2004; Erdirencelebi and Yalpir 2011; González et al. 2009). In the recent past, soft computing tools such as

✉ D. S. Manu
manu.cv13f05@nitk.edu.in

Arun Kumar Thalla
thallpce@gmail.com

¹ Department of Civil Engineering, National Institute of Technology Karnataka, Mangalore, Karnataka 575025, India

artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS) have also been widely used for wastewater treatment prediction studies (Belanche et al. 1998; Elmolla et al. 2010; Cakmakci 2007).

Nitrogen is a major wastewater nutrient and exists in various forms, including free ammonia, organic nitrogen, nitrate and nitrite each of which may be assessed for in a variety of ways. Fresh wastewater nitrogen is generally present in the ammonia and organic nitrogen forms, with the minute corpus of nitrite and nitrate forms (Sharma and Chopra 2015). The effluent may consist of either ammonia or nitrate nitrogen depending on the extent of nitrification, which exists within the treatment plant. Under routine conditions, the nitrite form of nitrogen does not exist in fat quantities due to its instantaneous oxidation or transformation to nitrate (Zhang and Gao 2000). Total Kjeldahl Nitrogen (TKN) is a chemical analysis to ascertain both the organic and the ammonia nitrogen. The TKN value corresponds to a total nitrogen concentration, which is the summation of organic nitrogen compounds and ammonia nitrogen [$\text{TKN} = \text{org-N} + \text{NH}_4\text{-N}$ (mg/L)]. Nitrogen mainly occurs in wastewater in the TKN form. After biological wastewater treatment, TKN mostly appears as oxidized nitrite (Liu et al. 2013).

The objective of the current study is to investigate the applicability of support vector machine (SVM) and adaptive neuro-fuzzy inference system (ANFIS) modeling approach for predicting the Kjeldahl Nitrogen removal from a domestic WWTP. Support vector machine is a unique state-of-the-art classification and regression technique based on the framework of Vapnik's statistical learning theory (Cortes and Vapnik 1995) designed to solve complex regression problems. The hybrid neuro-fuzzy approach developed from the combination of neural network and fuzzy system paves way for implementing an effective tool/algorithm for solving non-linear and complex real-world problems. Due to its abilities, such as handling imprecisions, uncertainties and large data sets, adaptive neuro-fuzzy inference system (ANFIS) is evolved to be one of the commonly used techniques. ANFIS trains the influencing parameters of the fuzzy inference system through a learning algorithm deduced from neural network (Jang 1993). Considering the difficulties associated with the conventional or analytical approaches and the experimentation/computational cost, SVM and ANFIS techniques are suitable choices to predict the Kjeldahl Nitrogen removal in the system.

Description of WWTP and data analysis

The data sets were obtained from the Kavoor Wastewater Treatment Plant (WWTP) situated at Mangalore which serves a population of 440,000. The design capacity of the

WWTP is 43.5 MLD, respectively. The normal operating DO in the aerobic reactor was about 1.7–2.5 mg/L. The sludge retention time was about 8–10 days with a hydraulic retention time of 7–8 h. The mixed liquor suspended solids (MLSS) maintained in the aerobic reactor was about 4200–4500 mg/L. The data set contains daily time series data analyzed and recorded at the WWTP plant during the period June–Sept 2014 with a total of 88 data points (period of 4 months) of seven variables, namely pH, total solids (TS), chemical oxygen demand (COD), temperature (T), free ammonia (FA), ammonia nitrogen (AN) and total Kjeldahl Nitrogen (TKN). The Kavoor WWTP adopts a biological treatment process, which possess the capability to remove phosphorus and nitrogen simultaneously under anaerobic and aerobic environments. The Kavoor WWTP consists of screening, grit chamber, anaerobic, aerobic reactors and a secondary clarifier as shown in Fig. 1. Complete removal of total Kjeldahl Nitrogen (TKN) is practically unachievable in the WWTP's having a pre-anaerobic system, wherein the anaerobic reactor is positioned behind the aerobic reactor and the mixed liquor involving nitrate is recirculated to the aerobic reactor from the secondary clarifier. The nitrate recirculation rate needs to be intensified, so as to improve the TKN removal efficiency, which steers to higher power consumption and dissolved oxygen (DO) return from the aerobic reactor (Liu et al. 2013).

The raw influent is fed into the bar screen, followed by grit chamber, anaerobic, and aerobic reactors, subsequently the sludge from the secondary clarifier is restored to the aerobic reactor. The treatment plant incorporates a simultaneous nitrification and denitrification (SND) process which initiates with partial nitrification of NH_4^+ to nitrite and successively continues with an immediate reduction of nitrite to N_2 gas. In SND process, nitrification and denitrification exist simultaneously in the same reactor basin under identical operating conditions (Breisha 2010). The main factors affecting nitrogen removal efficiency are temperature, nitrate concentration, dissolved oxygen, alkalinity, pH, BOD, COD and free ammonia concentration. At high temperatures (between 28 and 38 °C) the specific growth rate of ammonia oxidizing bacteria (AOB) will be higher than that of nitrite oxidizing bacteria NOB effecting in enhanced nitrogen removal rate via nitrite. Nitrifiers are vulnerable to temperature than heterotrophic bacteria. Optimal pH for effective nitrification is somewhere between 7 and 8.5. pH lower than 6 can cause inhibition. Alkalinity acts as a source of carbon for nitrifier growth. Nitrifiers are very sensitive to diverse kinds of compounds present in wastewater and get inhibited at very low DO levels. If the operating solids retention time (SRT) is lesser than the minimum SRT, nitrification process will be hampered. COD plays a role during denitrification



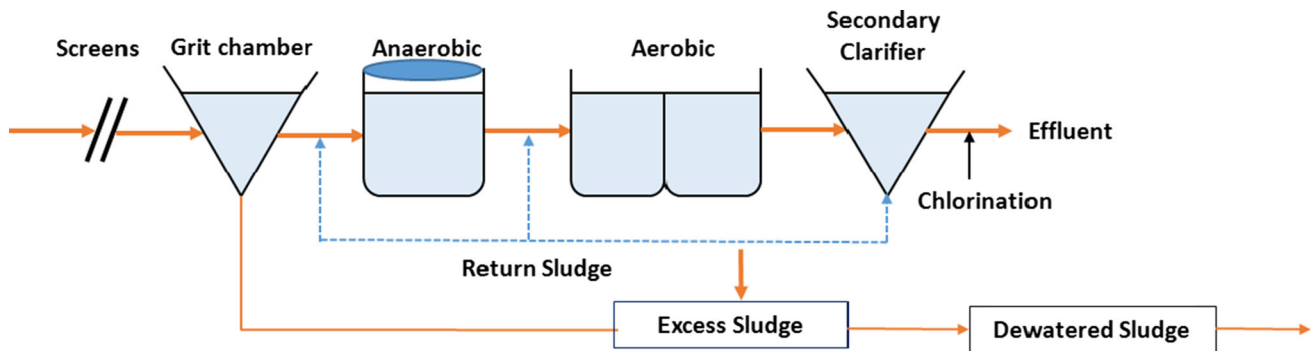


Fig. 1 Schematic flow diagram of Kavoor WWTP

process. Even though high DO concentrations are required to augment the activity of nitrifiers in the reactor, denitrification gets inhibited by excess oxygen. Free ammonia also inhibits the ammonium and nitrite oxidation during nitrification and denitrification processes. Hence, in the present context, the factors such as influent pH, COD, total solids (TS), temperature (T), free ammonia (FA), ammonia nitrogen (AN) and total Kjeldahl Nitrogen (TKN) are used as predictors to predict the effluent total Kjeldahl Nitrogen (TKN) concentrations using artificial intelligence (AI) models. The influent and effluent wastewater characteristics are analyzed on a daily basis by adopting the grab sampling technique. The details of sampling source and the laboratory methods of wastewater analysis are provided in Table 1. Sampling is carried out between 8 AM and 10 AM every day as the plant receives its peak flow. The descriptive statistics of the observed variables of WWTP are presented in Table 2. The X_{\max} , X_{\min} , X_{mean} , SD, & C_v denotes the maximum, minimum, mean, standard deviation and variance of the data respectively.

Methodology

Support vector machine

Support vector machine is a unique state-of-the-art classification and regression technique based on the framework of Vapnik's statistical learning theory (Cortes and Vapnik 1995) designed to solve complex regression problems. The SVM technique has been effectively used to perform multivariate function estimation, nonlinear regression problems, etc. due to its competence to escape from local minima, improved generalization capability and sparse representation of the solution (Vapnik 1999). SVM is based on structural risk minimization principle wherein it addresses the problem of overfitting by balancing the model's complexity. Non-linear problems are tackled by transforming them into linear ones in multi-dimensional feature space using Kernel functions. The structure

of SVM is as represented in Fig. 2. With the innovation of Vapnik's ϵ -insensitivity loss function, the SVM is still more capable to solve nonlinear regression problems (Smola and Schölkopf 2004). In order to achieve a good generalization performance, it is essential to find certain optimal hyper-parameters of SVM model. The hyper-parameters that need to be tuned are the regularization parameter (C) that controls the generalization performance of SVM, secondly the kernel parameter specific to the type of kernel adopted and finally the radius of ϵ —insensitive zone which determines the number of support vectors (Cristianini and Shawe-Taylor 2000; Kecman 2001). A brief description and derivation of support vector regression can be referred from various literatures (Smola and Schölkopf 2004; Cristianini and Shawe-Taylor 2000; Raghavendra and Deka 2015a).

ANFIS architecture

ANFIS, a hybrid fuzzy logic-based technique integrated with the learning power of artificial neural network improves the performance of any kind of intelligent system by utilizing knowledge acquired after learning. For a real-time input–output dataset, a hybrid learning algorithm such as ANFIS constructs a backpropagation gradient descent and least squares methods associatively to frame a fuzzy inference system whose membership function parameters are iteratively tuned or adjusted. Adaptive neuro-fuzzy inference systems comprise of a mainly five layers—rule base, database, fuzzification interface, defuzzification interface and decision making unit (Jovanovic et al. 2004; Raghavendra and Deka 2015b). The generalized ANFIS architecture proposed is summarized below.

The ANFIS is a fuzzy Sugeno model that allocates the structure of adaptive systems to assist learning and adaptation. ANFIS architecture comprises of five layers. Every single node in layer 1 is an adaptive node with a node function which may be anyone among the membership functions. Every node of layer 2 is a fixed node labeled ' π ' which signposts the firing strength of each rule. All nodes

Table 1 Sampling source and the laboratory methods of wastewater analysis

Characteristic	Sampling source	Test method
pH	Influent and effluent	IS 3025 (Part 11): 1983 (RA 2006) Potentiometric method
TS (mg/L)		IS 3025 (Part 15): 1984a (RA 2003) Gravimetric method
COD (mg/L)		IS 3025 (Part 58): 2006 Open reflex method
T (°C)		IS 3025(Part 9): 1984b (RA 2002) Mercury –in-glass thermo meter method
FA (mg/L)		IS 3025 (Part 34): 1988 (RA 2009) Macro-Kjeldahl method with calorimetric analysis
AN (mg/L)		IS 3025 (Part 34): 1988 (RA 2009) Spectrophotometric method
TKN (mg/L)		IS 3025 (Part 34): 1988 (RA 2009) TKN distillation method

TS total solids, COD chemical oxygen demand, T temperature, FA free ammonia, AN ammonia nitrogen, TKN total Kjeldahl Nitrogen

of layer 3 are fixed nodes labeled as ‘N’ which demonstrates the normalized firing strength of each rule. The layer 4 is as similar to layer 1 wherein every node is an adaptive node governed by a node function. The layer 5

being a single fixed node labeled ‘Σ’, representing the final output (f), defined as the summation of all arriving signals.

Figure 2, shows the implementation of two fuzzy rules using ANFIS architecture. The appropriate choice of the type and the parameters of the fuzzy membership functions and rules play a vital role in achieving the desired performance but in most circumstances, it is problematic (Raghavendra et al. 2015). Sometimes these parameters are chosen on the basis of trial and error method which enlightens the importance of tuning the fuzzy system. The main objective of training the ANFIS system is to govern the optimal premise and resultant parameters. ANFIS can be used to train the FIS model by modifying the membership function parameters based on error chosen criterion to cope with the training data. The FIS model having parameters related to the least checking data model error is selected, when ANFIS contains the checking data and training data.

Performance evaluation

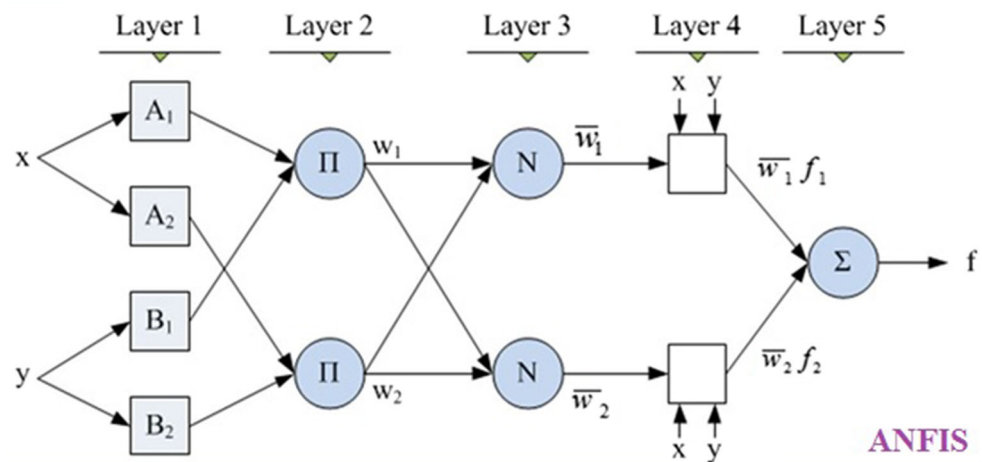
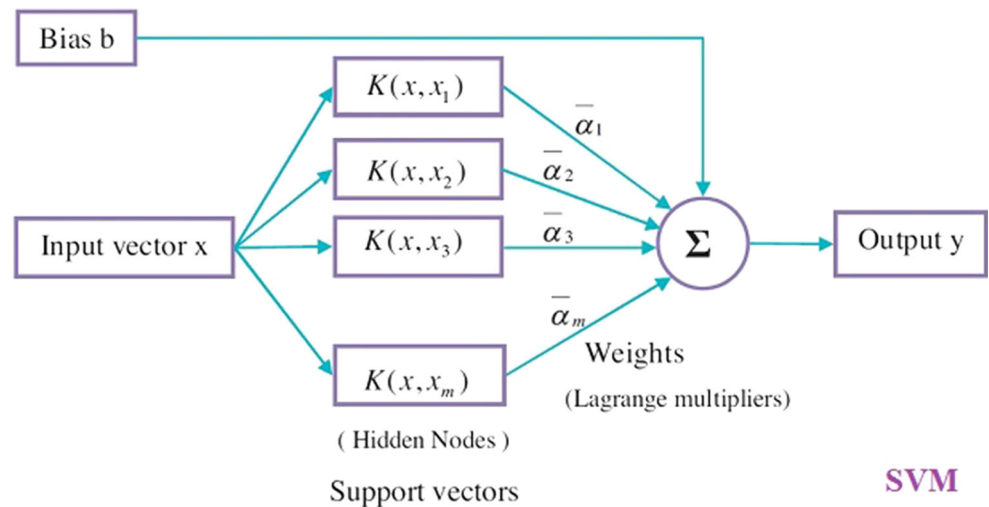
The level of confidence over the predictions of any developed model is assessed by using suitable statistical indices. Correlation coefficient (CC), root mean square error (RMSE) and Nash–Sutcliffe error (NSE) were used to evaluate the model accuracies. Although RMSE values are used to distinguish model performance in training and testing period, it can also be used to compare the performance of individual model to other predictive models. To

Table 2 Statistical indices of various parameters of WWTP

Parameters		Statistical indices				
		X _{max}	X _{min}	X _{mean}	S _d	C _v
Train phase						
Predictors	Influent pH	6.70	6.30	6.45	0.09	0.0078
	Influent TS (mg/L)	670.00	367.00	487.05	56.84	3230.64
	Influent COD (mg/L)	592.00	264.00	389.72	73.14	5349.11
	Influent T (°C)	34.00	27.00	29.29	1.60	2.55
	Influent FA (mg/L)	0.16	0.05	0.09	0.025	0.0006
	Influent AN (mg/L)	29.00	10.00	17.12	4.40	19.33
	Influent TKN (mg/L)	37.00	16.00	23.71	5.35	28.59
Predictand	Effluent TKN (mg/L)	32.00	11.00	19.72	5.00	24.98
Test phase						
Predictors	Influent pH	6.70	6.30	6.4478	0.1039	0.0108
	Influent TS (mg/L)	626.00	382.00	460.78	51.68	2671.27
	Influent COD (mg/L)	504.00	200.00	329.39	84.87	7202.34
	Influent T (°C)	30.00	27.00	28.4783	0.6653	0.4427
	Influent FA (mg/L)	0.08	0.02	0.0436	0.0126	0.0002
	Influent AN (mg/L)	11.00	5.00	8.45	1.53	2.3550
	Influent TKN (mg/L)	16.00	9.00	13.13	1.96	3.8458
Predictand	Effluent TKN (mg/L)	14.00	7.00	10.83	1.80	3.2411



Fig. 2 Schematic diagram of SVM and ANFIS Structure (Source: Raghavendra and Deka 2014)



assess the performance of ANFIS models the following statistical indices were adopted.

1. Correlation coefficient (CC)

$$CC = \frac{\sum_{i=1}^n (X_i - \bar{X}) \times (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \times \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

2. Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i - Y_i)^2}{n}} \quad (2)$$

3. Nash–Sutcliffe coefficient (NSE)

$$NSE = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (3)$$

where X = observed/actual values; Y = predicted values; \bar{X} = mean of actual data values; n = total number of values.

Results and discussion

The dataset is split into ‘train dataset’ which includes 74% (65 data points) of data in the period of 2nd June 2014 to 30th August 2014 and ‘test dataset’ composed of the remaining 26% data (23 data points) in the period of 1st September 2014 to 30th September 2014. The train dataset was used to build/simulate the model and the test dataset was employed to evaluate the performance of the built model. In order to investigate the dependency between variables that influence total Kjeldahl Nitrogen (TKN), cross-correlation coefficients between effluent TKN and each input parameter were analyzed and are presented in Table 3. This data were exercised to assist in selecting input variables for ANFIS and SVM models. From

Table 3 Cross-correlation between effluent total Kjeldahl Nitrogen (TKN) and other parameters

Parameter	Effluent total Kjeldahl Nitrogen (TKN)	
	Train data	Test data
Influent pH	−0.597	−0.532
Influent TS (mg/L)	0.654	0.628
Influent COD (mg/L)	0.723	0.698
Influent T (°C)	0.646	0.622
Influent FA (mg/L)	0.872	0.765
Influent AN (mg/L)	0.916	0.853
Influent TKN (mg/L)	0.952	0.920

Table 3, it can be noticed that the effluent TKN at the time (t) is strongly correlated with the influent total Kjeldahl Nitrogen concentration [with a correlation value of 0.952(in train dataset) and 0.92(test)]; Influent ammonia nitrogen concentration [with a correlation value of 0.916(train) and 0.85(test)]; and the influent free ammonia concentration [with a correlation value of 0.87(train) and 0.76(test)]. The cross-correlation coefficients between the effluent TKN and other variables (influent total solids, COD concentrations, temperature) were also found to be fairly influential. The cross-correlation coefficients between the effluent TKN and the influent pH ranged from −0.597(train) and −0.532(test). The negative correlation indicates that a high occurrence or amount of TKN is rendered in the effluent during decreased pH of the influent.

The analysis is carried out to predict the concentration of effluent Kjeldahl Nitrogen using influent pH, TS, COD, Free ammonia, ammonia nitrogen, Kjeldahl Nitrogen as

input variables. The cross-validation search is used to determine the optimal SVM hyper-parameters (C , γ and ϵ). SVM with RBF kernel function is implemented in the present case. The optimal parameters obtained after tuning the SVM model are as tabulated in Table 4. The modeling of ANFIS is carried out in MATLAB platform. The results obtained from SVM and ANFIS models with Gbell and trapezoidal MFs are depicted in the form of various statistical indices like RMSE, CC and NSE through tables and various plots. The optimal ANFIS architecture as presented in Table 4 is obtained after tuning fuzzy MF and rules of certain number and type.

The prediction errors of the models in the training and testing phases are as presented in Table 5. In the SVM model, the RMSE and NSE are significantly less in both training and testing stages when compared to that of ANFIS models. The magnitude of RMSE and NSE computation infers that the ANFIS model with Gbell membership function closely predicts the effluent Kjeldahl Nitrogen concentration than that of trapezoidal membership function. Here, the $RMSE = 0.795$ mg/L, $NSE = 0.79$ and $CC = 0.85$ of ANFIS model with Gbell membership function during test phase verifies the close agreement of concentration of effluent Kjeldahl Nitrogen with the observed concentration. The comparative evaluation of results obtained from Gbell and trapezoidal ANFIS models along with the SVM model during the prediction of effluent Kjeldahl Nitrogen is as presented in the form of graph (Fig. 3).

The SVM algorithm outperformed the ANFIS models, particularly in the testing stage. The prediction errors and correlation statistic of the SVM algorithm is relatively

Table 4 Details of SVM and ANFIS architecture

ANFIS architecture		SVM model	
No. of membership function (MF)	3	Optimal C	15
Algorithm selected	Hybrid	Optimal ϵ	0.15
No. of Epoch given	500	Optimal γ	5
FIS generated	Grid partition	nsv	47%
No. of membership (MF) type	Constant		
Member ship function (MF) used	Gbell & trapezoidal		

Table 5 Statistical results of SVM and ANFIS models

Statistical indices	ANFIS models				SVM model	
	GBELL MF		Trapezoidal MF			
	Train	Test	Train	Test	Train	Test
CC	0.97	0.85	0.96	0.79	0.98	0.91
RMSE (mg/L)	0.198	0.795	0.532	1.104	0.155	0.232
NSE	0.96	0.79	0.97	0.58	0.98	0.85

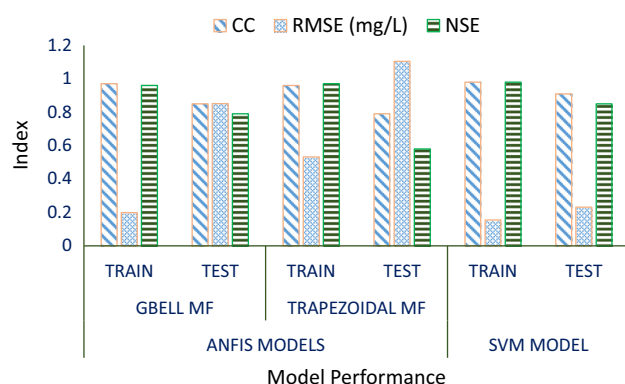


Fig. 3 Comparative performance evaluation of SVM and ANFIS models

better than the ANFIS models as presented in Figs. 3 and 4, respectively. It is common to see that each and every model gets better solutions in the training stage as compared to that of testing stage. The possible reason for this is, the models will be trained over the range of dataset with specific maximum and minimum values. The mean of the dataset will also influence during training of a model. However, during testing of the model with another dataset of different minima and maxima, the model is usually unsuccessful to catch up the limits of the testing dataset. From the time series graph as presented in Fig. 4 during the effluent Kjeldahl Nitrogen prediction, it is observed that the SVM model closely follows the observed time series. The ANFIS model with Gbell MF appears to have the accepted accuracy during both training and testing phase.

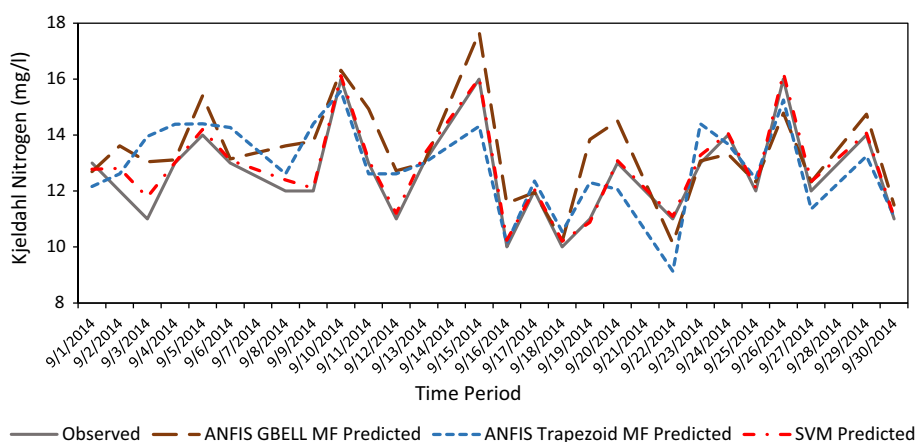
Figure 5 shows closely spaced scatters of the predicted and observed effluent Kjeldahl Nitrogen concentrations of SVM and ANFIS models during the testing phase. The reasonable dependence of a variable can be verified through the coefficient of determination (R^2) which ranges between 0 and 1 signposting the predictable extent of the dependent variable. The data points in the upper and lower extremes of the scatter plot of SVM model do not deviate

to a great extent from the line of best fit indicating the goodness of the fit/model. In SVM model 82.48% of the variations in total Kjeldahl Nitrogen prediction is explained by taking into account of pH, TS, COD, T, FA and AN as predictors. It can be observed that ANFIS model with trapezoidal MF has more number of outliers than that of the SVM and Gbell ANFIS models during the test phase. From this, it can be ascertained that SVM model has higher consistency and robust performance during prediction.

Summary and conclusions

Much research has endorsed that biological wastewater treatment is an extremely viable treatment technology regarding nitrification–denitrification and phosphorus removal. In conjunction with optimized plant design and operating parameters, the biological wastewater treatment guarantees high effluent quality in terms of nitrates, ammonia, and phosphates existing in wastewater. According to contemporary European regulation, the total phosphorus and nitrogen in treated effluent should be in the range of 1–2 and 10–15 mg/L, respectively. In many situations, where the risk of public exposure to the reclaimed water exists, effective monitoring of effluent quality is necessary. The data related to influent pollutants, including the total suspended solids (TSS) and COD are utilized for immediate or short-term effluent quality prediction to provide information for efficient operation of the treatment process. In this study, the artificial intelligence models—SVM and ANFIS are being applied for the prediction of effluent Kjeldahl Nitrogen concentration yielded from a biological wastewater treatment plant. SVM and ANFIS models with Gbell and trapezoidal membership functions are tested in the study with input variables such as influent pH, TS, COD, Free ammonia, ammonia nitrogen and Kjeldahl Nitrogen. From the results presented above, cross-validation search was able to set the SVM parameters

Fig. 4 Predicted effluent Kjeldahl Nitrogen concentration of SVM and ANFIS models during testing



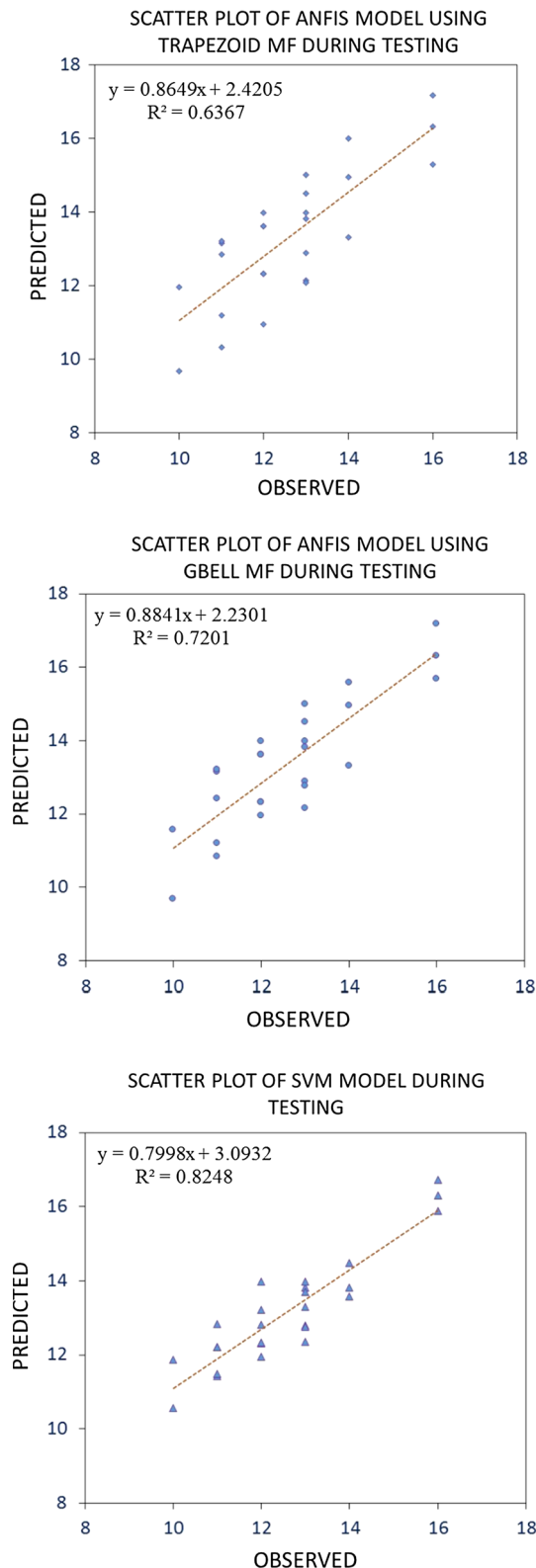


Fig. 5 Scatter plot of observed v/s Predicted of SVM and ANFIS models during testing

efficiently and thereby improve the forecasting efficiency of SVM. SVM models provided reliable prediction results than the ANFIS models. Among ANFIS models, Gbell MF MODEL was found to be slightly efficient in modeling the nonlinear time series. However, due to the computational complexity of various membership functions, trapezoidal membership function was found to be incompatible to model the effluent Kjeldahl Nitrogen concentration in the present study.

Acknowledgements The authors would like to thank the Mangalore City Corporation, Dakshina Kannada District, Karnataka for providing the necessary data required for research and the Department of Civil Engineering, National Institute of Technology Karnataka for the necessary infrastructural support.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Belanche LA, Valdés J, Lluís AB, Valdés J, Comas J, Ignasi R, and Roda MP (1998) Modeling the input–output behaviour of wastewater treatment plants using soft computing techniques. In: Proceedings of BESAI'98. Binding environmental sciences and AI. Workshop held as part of ECAI'98: European Conference on Artificial Intelligence. Brighton, UK
- Boelee NC, Temmink H, Janssen M, Buisman CJN, Wijffels RH (2011) Nitrogen and phosphorus removal from municipal wastewater effluent using microalgal biofilms. *Water Res* 45:5925–5933. doi:[10.1016/j.watres.2011.08.044](https://doi.org/10.1016/j.watres.2011.08.044)
- Breisha GZ (2010) Bio-removal of nitrogen from wastewaters—a review. *Nat Sci* 8(12):210–228
- Cakmakci M (2007) Adaptive neuro-fuzzy modelling of anaerobic digestion of primary sedimentation sludge. *Bioprocess Biosyst Eng* 30:349–357. doi:[10.1007/s00449-007-0131-2](https://doi.org/10.1007/s00449-007-0131-2)
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. doi:[10.1007/BF00994018](https://doi.org/10.1007/BF00994018)
- Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, New York, USA
- Denai MA, Palis F, Zeghib A (2004) ANFIS based modelling and control of non-linear systems: a tutorial. In: 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583):4. doi:[10.1109/ICSMC.2004.1400873](https://doi.org/10.1109/ICSMC.2004.1400873)
- Elmolla ES, Chaudhuri M, Eltoukhy MM (2010) The use of artificial neural network (ANN) for modeling of COD removal from antibiotic aqueous solution by the Fenton process. *J Hazard Mater* 179:127–134. doi:[10.1016/j.jhazmat.2010.02.068](https://doi.org/10.1016/j.jhazmat.2010.02.068)
- Erdirencelebi D, Yalpir S (2011) Adaptive network fuzzy inference system modeling for the input selection and prediction of anaerobic digestion effluent quality. *Appl Math Model* 35:3821–3832. doi:[10.1016/j.apm.2011.02.015](https://doi.org/10.1016/j.apm.2011.02.015)

- González C, García PA, Muñoz R (2009) Effect of feed characteristics on the organic matter, nitrogen and phosphorus removal in an activated sludge system treating piggery slurry. *Water Sci Technol: J Int Assoc Water Pollut Res* 60:2145–2152. doi:10.2166/wst.2009.579
- Guo YM, Liu YG, Zeng GM, Hu XJ, Xu WH, Liu YQ, Huang HJ (2014) An integrated treatment of domestic wastewater using sequencing batch biofilm reactor combined with vertical flow constructed wetland and its artificial neural network simulation study. *Ecol Eng* 64:18–26. doi:10.1016/j.ecoleng.2013.12.040
- IS 3025-11 (1983): Methods of sampling and test (physical and chemical) for water and wastewater, Part 11: pH value [CHD 32: Environmental Protection and Waste Management] <https://law.resource.org/pub/in/bis/S02/is.3025.11.1983.pdf>. Accessed 3 Nov 2016
- IS 3025-9 (1984): Methods of sampling and test (physical and chemical) for water and wastewater, Part 9: Temperature [CHD 32: Environmental Protection and Waste Management] <https://law.resource.org/pub/in/bis/S02/is.3025.09.1984.pdf>. Accessed 3 Nov 2016
- IS 3025-15 (1984): Methods of sampling and test (physical and chemical) for water and wastewater, Part 15: Total residue (total solids-dissolved and suspended) [CHD 32: Environmental Protection and Waste Management] <https://law.resource.org/pub/in/bis/S02/is.3025.15.1984.pdf>. Accessed 3 Nov 2016
- IS 3025-34 (1988) Methods of sampling and test (physical and chemical) for water and wastewater, Part 34: Nitrogen [CHD 32: Environmental Protection and Waste Management] <https://law.resource.org/pub/in/bis/S02/is.3025.34.1988.pdf>. Accessed 3 Nov 2016
- IS 3025-58 (2006): Methods of sampling and test (physical and chemical) for water and wastewater, Part 58: Chemical oxygen demand (COD) [CHD 32: Environmental Protection and Waste Management] <https://law.resource.org/pub/in/bis/S02/is.3025.58.2006.pdf>. Accessed 3 Nov 2016
- Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23:665–685. doi:10.1109/21.256541
- Jovanovic BB, Reljin IS, Reljin BD (2004) Modified ANFIS architecture—improving efficiency of ANFIS technique. In: 7th Seminar on Neural Network Applications in Electrical Engineering, 2004. NEUREL 2004. doi:10.1109/NEUREL.2004.1416577
- Kecman V (2001) Learning and soft computing: support vector machines, neural networks and fuzzy logic models. MIT Press, Cambridge MA, USA
- Kumar TA, Saravanan S (2009) Treatability studies of textile wastewater on an aerobic fluidized bed biofilm reactor (FABR): a case study. *Water Sci Technol* 59:1817–1821. doi:10.2166/wst.2009.207
- Liu G, Xu X, Zhu L, Xing S, Chen J (2013) Biological nutrient removal in a continuous anaerobic–aerobic–anoxic process treating synthetic domestic wastewater. *Chem Eng J* 225:223–229. doi:10.1016/j.cej.2013.01.098
- Miller RM, Itoyama K, Uda A, Takada H, Bhat N (1997) Modeling and control of a chemical waste water treatment plant. *Comput Chem Eng* 21:S947–S952. doi:10.1016/S0098-1354(97)87624-7
- Nair VV, Dhar H, Kumar S, Thalla AK, Mukherjee S, Wong JW (2016) Artificial neural network based modeling to evaluate methane yield from biogas in a laboratory-scale anaerobic bioreactor. *Bioresour Technol* 217:90–99. doi:10.1016/j.biortech.2016.03.046
- Pai TY, Yang PY, Wang SC, Lo MH, Chiang CF, Kuo JL, Chang YH (2011) Predicting effluent from the wastewater treatment plant of industrial park based on fuzzy network and influent quality. *Appl Math Model* 35:3674–3684. doi:10.1016/j.apm.2011.01.019
- Ráduly B, Gernaey KV, Capodaglio AG, Mikkelsen PS, Henze M (2007) Artificial neural networks for rapid WWTP performance evaluation: methodology and case study. *Environ Model Softw* 22:1208–1216. doi:10.1016/j.envsoft.2006.07.003
- Raghavendra NS, Deka PC (2014) Support vector machine applications in the field of hydrology: a review. *Appl Soft Comput* 19:372–386. doi:10.1016/j.asoc.2014.02.002
- Raghavendra NS, Deka PC (2015a) Forecasting monthly groundwater level fluctuations in coastal aquifers using hybrid Wavelet packet–Support vector regression. *Cogent Eng* 2(1):999414. doi:10.1080/23311916.2014.999414
- Raghavendra NS, Deka PC (2015b) Multistep ahead groundwater level time-series forecasting using gaussian process regression and ANFIS. In: advanced computing and systems for security. Springer, India, pp 289–302. doi: 10.1007/978-81-322-2653-6_19
- Raghavendra NS, Sudheer C, Deka PC (2015) Genetic algorithm optimized support vector regression model for forecasting groundwater level time-series. In: 20th International Conference on Hydraulics, Water Resources and River Engineering (HYDRO 2015 International), IIT, Roorkee
- Sharma AK, Chopra AK (2015) Removal of nitrate and sulphate from biologically treated municipal wastewater by electrocoagulation. *Appl Water Sci*. doi:10.1007/s13201-015-0320-0
- Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14(3):199–222. doi:10.1023/B:STCO.0000035301.49549.88
- Thalla AK, Bhargava R, Kumar P (2010) Nitrification kinetics of activated sludge-biofilm system: a mathematical model. *Bioresour Technol* 101:5827–5835. doi:10.1016/j.biortech.2010.03.014
- Vapnik VN (1999) An overview of statistical learning theory. *IEEE Trans Neural Netw Publ IEEE Neural Netw Counc* 10(5):988–999. doi:10.1109/72.788640
- Zhang B, Gao T (2000) An anoxic/anaerobic/aerobic process for the removal of nitrogen and phosphorus from wastewater. *J Environ Sci Health Part A* 35(10):1797–1801. doi:10.1080/10934520009377075